



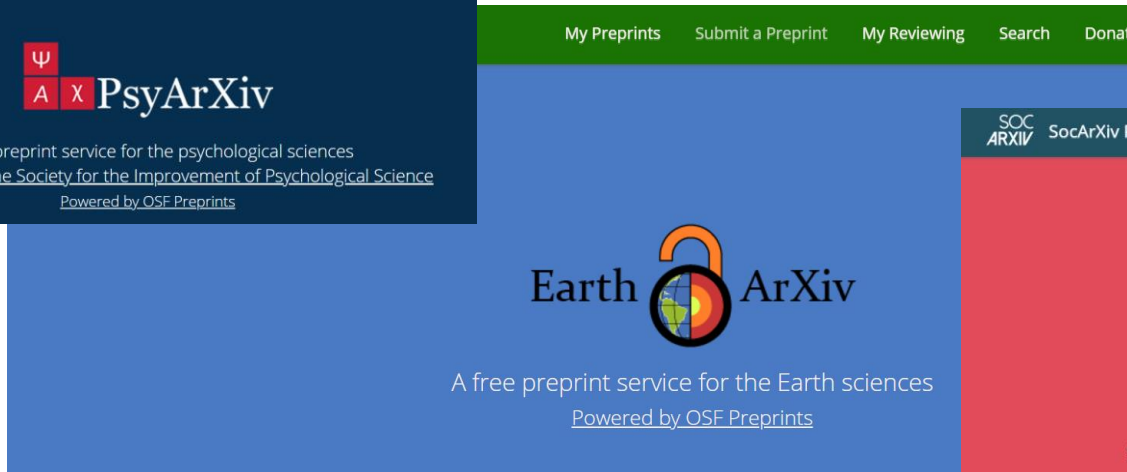
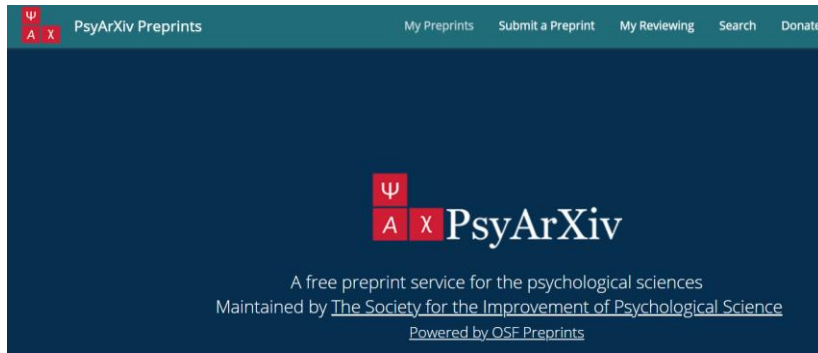
Acceptance and Use of Preprints in Diverse Domains and Disciplines

Tom Narock and Evan Goldstein

Center for Open Science

- The results presented today are from a study with my colleague Evan Goldstein
- Our study was focused on preprint systems offered by the Center for Open Science (COS)
 - COS is a non-profit organization based in Charlottesville, VA
 - Their mission is “to increase openness, integrity, and reproducibility of research”
- COS created the Open Science Framework (OSF)
 - OSF is a free and open source project management tool that supports researchers
- In 2016, COS used OSF to launch a shared preprint infrastructure

OSF Preprints



Shared Hardware and Software

- shared storage for preprints and related resources (data, software, etc.)
- shared website hosting
- customized view for your community

What's in a Name?

- At time of our study, COS hosted 23 preprint systems sharing a common infrastructure
 - Reduces costs
 - Helps get new communities interested in preprints up and running quickly
- COS preprint systems either represent a domain or a geographical area
 - EarthArXiv = preprints for the Earth sciences
 - Ina-Rxiv = preprints for the Indonesian research community
- The names are variants of “arXiv” in honor of one of the original preprint systems

Open Access to All Preprints



- We wrote software to query all COS preprint systems
- We wanted to understand
 - Differences in usage amongst domains
 - Usage of preprints vs. postprints
 - Topic overlap across domains
 - Open access to peer reviewed version of manuscript
 - Connections amongst authors

Basic Criteria

- We required
 1. At least 100 total submissions
 2. English language manuscripts
 3. Manuscripts must be accessible through OSF Application Programming Interface
 4. Service must have a domain/community focus
 - Excludes general purpose services like OSF Preprints and Thesis Commons
- **9 services met all the criteria**

Basic Metrics

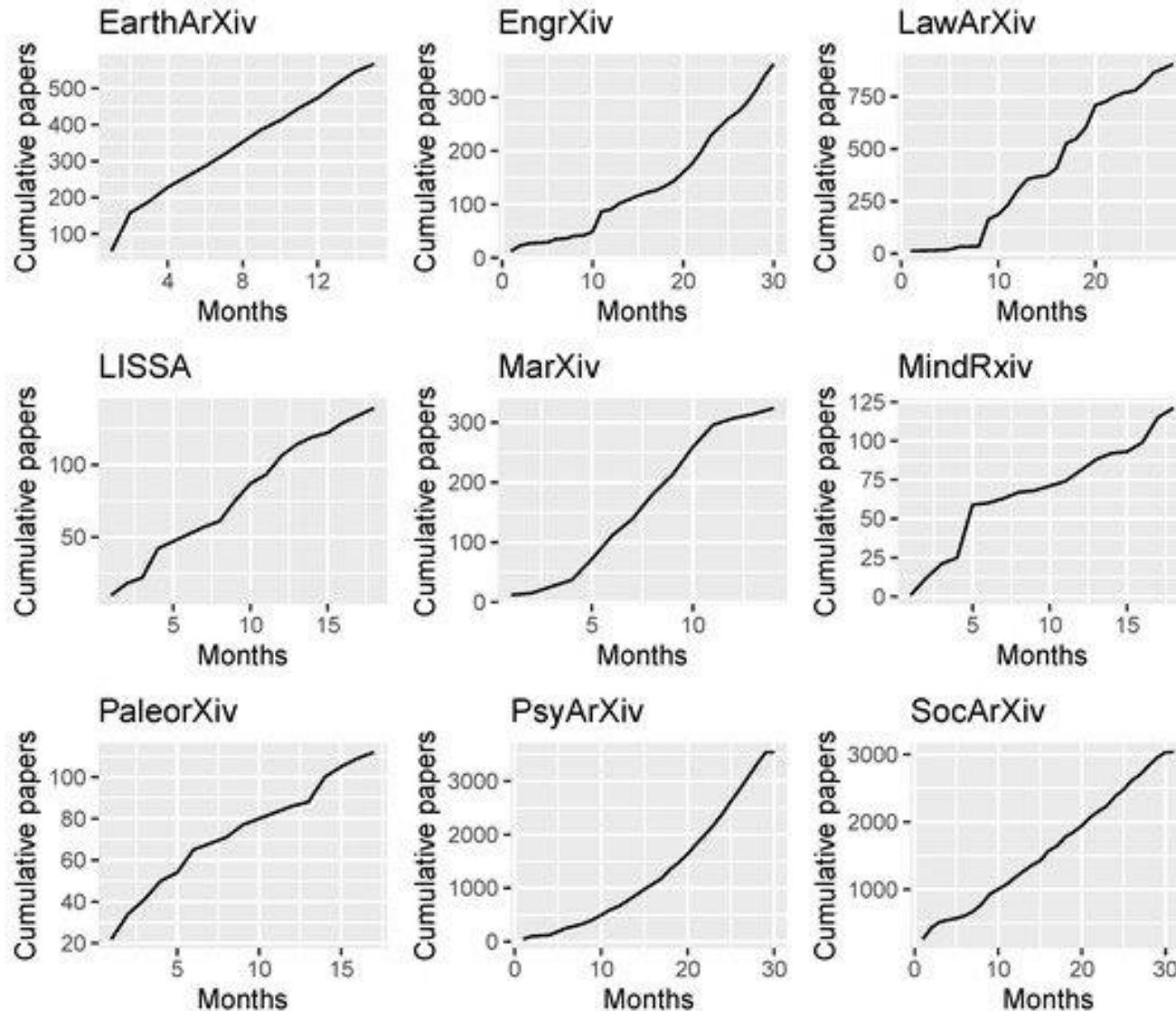
Preprint System	Domain	Total Papers	Total Authors	Distinct Authors
PsyArXiv	Psychology	3534	12,439	7342
SocArXiv	Sociology	3034	5337	3017
LawArXiv	Law	905	1186	463
EarthArXiv	Earth and Planetary Science	567	2353	1659
EngrXiv	Engineering	362	961	664
MarXiv	Marine Science	324	960	627
LISSA	Library and Information Science	139	257	175
MindRxiv	Mind and Contemplative Practices	121	413	285
PaleorXiv	Paleontology	112	343	236

The Distinct Authors number is a proxy for the breadth of a service.

A low percentage of distinct authors is indicative of a small group of researchers repeatedly submitting to the service. Conversely, a high percentage of distinct authors indicates more uptake of the service.

Distinct authors values cluster between 57–69% with LawArXiv an outlier at the low end with 39% and EarthArXiv the highest at 71%.

Basic Metrics



- 8 domains are growing linearly
- The psychological sciences community is growing superlinearly

Preprints vs. Postprints

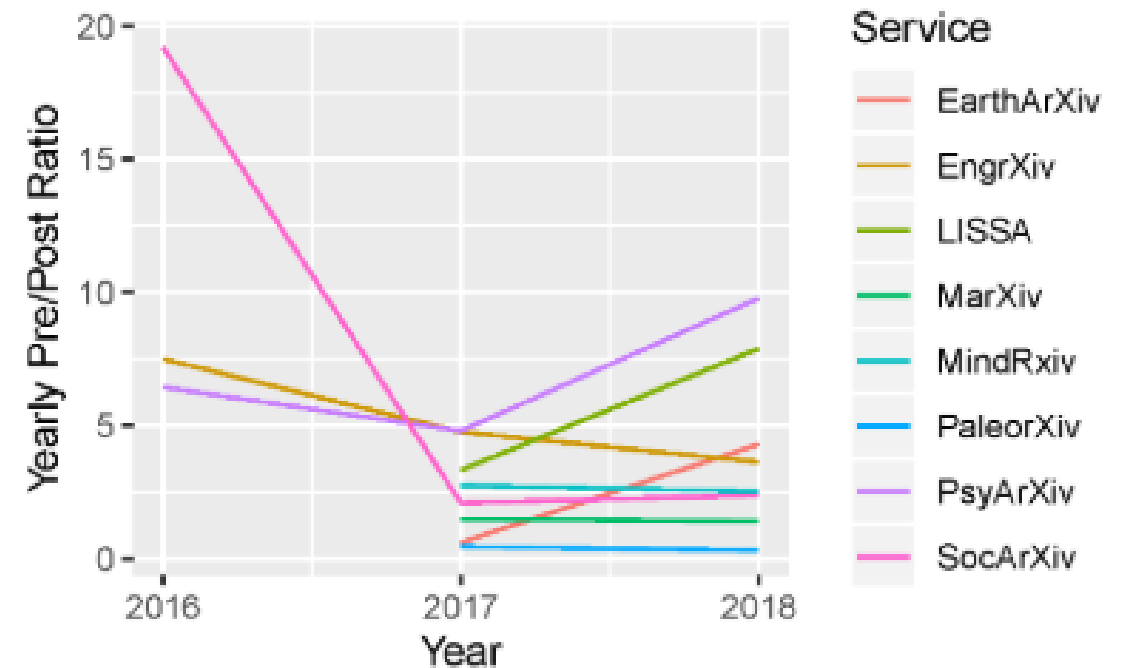
- COS preprint systems accept a variety of submissions
 - Preprints – non-peer reviewed manuscripts
 - Postprints – author's version of peer-reviewed manuscript, submitted to provide free access
 - Software
 - Data and related research materials
- We were interested in knowing the ratio of preprints to postprints
- Are COS preprint systems being used as preprint systems or as repositories for already published work?

Preprints vs. Postprints

- COS preprint systems invite authors to return after their manuscript appears in a peer-reviewed journal
 - Authors can add the peer-reviewed DOI to the preprint metadata
- We used these peer-reviewed DOIs to determine data of journal publication
- This is compared to date of submission to COS system

Limitation

- postprint designation requires the peer-reviewed DOI to be entered by the author in the COS service



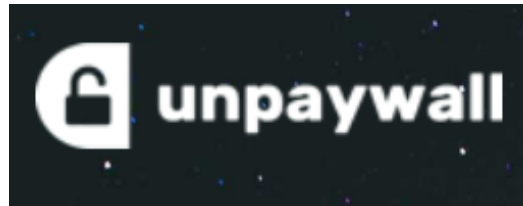
More on Limitations

- In order to estimate the extent to which we underestimate the number of COS manuscripts appearing in peer-reviewed journals, we chose 12 random preprints from each system
- We manually entered the titles into Google Scholar
- Found that 36% of the time a peer-reviewed article existed, but was not recorded in the COS system.
- While our 36% false-positive rate is less than what was found in bioRxiv [Abdill and Blekham, 2019], we caution that our sample size was smaller and the methodology varied

Abdill, R.J.; Blekham, R. Meta-Research: Tracking the popularity and outcomes of all bioRxiv preprints. *eLife* **2019**, *8*, e45133.

Open Access to Peer Reviewed Version?

- What about those peer-reviewed versions? Are they being published open access?
- The Unpaywall API returns information on which journal articles are published as Gold Open Access - in venues that are included in the Directory of Open Access Journals (DOAJ)



- Preprints that are eventually published in Gold Open Access venues varies widely.
 - From 0% (MarXiv, LISSA, and MindRxiv) to 33% (PaleoarXiv)
- Postprints are Gold Open Access before being deposited in Green Open Access COS systems
 - 39% of MindRxiv, 35% of SocArXiv, and 33% of LISSA postprints were published in Gold Open Access journals

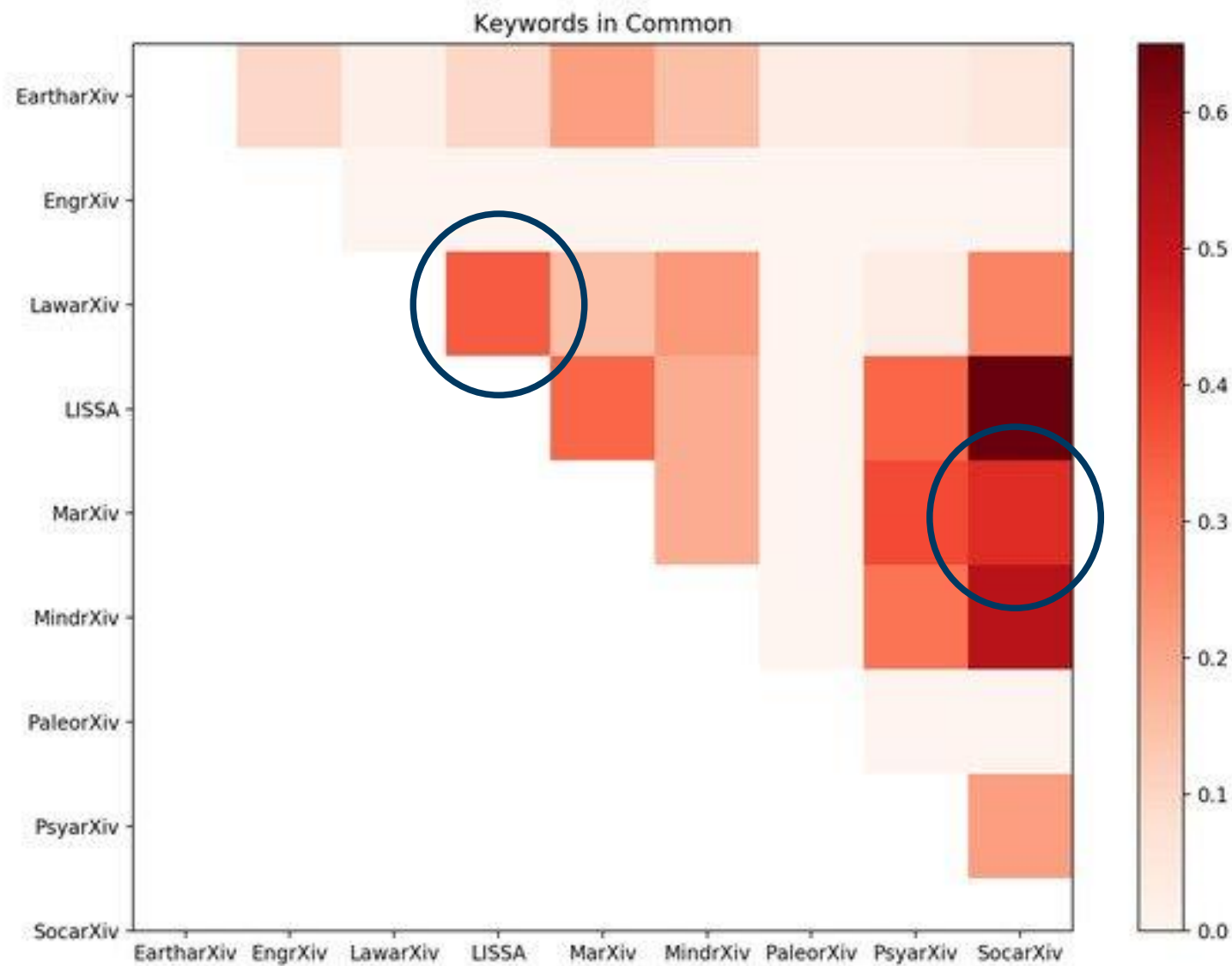
Overlap ranges from 0 to 1

0 = no overlap

1 = all topics in one system found in the other

Some was a surprise...

- Law community and Library Sciences
- Marine sciences and sociology community



Annotations

- COS uses Hypothes.is to enable commenting on preprints
- A total of 135 annotations were found across the >9000 papers on the nine services

Preprint System	Hypothes.is Annotations	Annotations per Manuscript
PsyArXiv	89	0.03
SocArXiv	14	0.01
LawArXiv	0	0.00
EarthArXiv	12	0.02
EngrXiv	2	0.01
MarXiv	3	0.01
LISSA	0	0.00
MindRxiv	1	0.01
PaleorXiv	14	0.13

Annotations

- COS uses Hypothes.is to enable commenting on preprints
- A total of 135 annotations were found across the >9000 papers on the nine services

Preprint System	Hypothes.is Annotations	Annotations per Manuscript
PsyArXiv	89	0.03
SocArXiv	14	0.01
LawArXiv	0	0.00
EarthArXiv	12	0.02
EngrXiv	2	0.01
MarXiv	3	0.01
LISSA	0	0.00
MindRxiv	1	0.01
PaleorXiv	14	0.13



PaleorXiv has 5x more annotations per manuscript as the next closest service

Wikipedia

- An additional place where preprints/postprints can provide value is in Wikipedia
- One of the largest websites in terms of global web traffic, an entry point to scholarly literature, and a site with page views that eclipse those of the primary literature.
- Yet, only 29% of the references in Wikipedia refer to free open access papers, and an additional 10% have a free open access version available but without an active link to the OA version.
- We examined the March 2018 release of Wikipedia references to look for COS preprint DOIs
- We found only 3.

Usage and Download Statistics

- On average, a COS preprint system gets a few hundred visits per month
- Visitors come from all over the world (we don't have usage data for all the systems)
- Individual paper downloads vary widely as you might expect
 - Some papers have a handful of downloads
 - Others have several hundred
- We caution that the download numbers do not have any quality control

Summary

- Variability amongst communities
 - Some communities are slow to accept preprints - struggle to get 100 papers
 - Other take off linearly
 - One exponential growth
- Some systems were excluded from our study, but are doing very well
 - The Indonesian preprint system has over 11,000 submissions
 - The Thesis Commons has over 600 undergraduate and graduate theses

Summary

- Our COS metrics can be placed in context by comparison with bioRxiv
 - 37,648 preprints uploaded to the biological preprint system
 - Their 2,100 papers per month is about the total number of preprints in our two largest domains
 - bioRxiv was founded in 2013 – three more years of community outreach
- What we find surprising is the high percentage of COS postprints that appear in DOAJ publications.
- It is unclear why, having already published Gold OA, authors would also publish at a preprint system.

Summary

- Given the overlap in topics, users would benefit from keyword alignment with non-COS preprint systems (e.g., bioRxiv and arXiv)
- 3 of the 9 systems studied show evidence of being preprint systems
- 5 of the 9 systems studied show evidence of trending towards having more postprints
- 1 system (LawArXiv) was excluded as the domain doesn't peer-review (much)
- Opportunities for preprints to reach a broader audience – e.g. Wikipedia – not taken advantage of

Reference

- This presentation is based on

Narock and Goldstein, 2019, Quantifying the Growth of Preprint Services Hosted by the Center for Open Science

<https://www.mdpi.com/2304-6775/7/2/44/htm#B15-publications-07-00044>